

# Develop and evaluate biomedical time series data clustering models based on Fluctuation Based Clustering (FBC)

Maximilian Koehler

2022-03-17

A proposal for a potential R package in the course of my master thesis at the chair for Computational Physiology and Biostatistics (UKBB/University of Basel) to perform conveniently Fluctuation Based Clustering (FBC) and corresponding evaluation procedures on medical time series data.

The following publications serve as the basis and main inspiration for the creation of the package: <https://pubmed.ncbi.nlm.nih.gov/28866644/> <https://pubmed.ncbi.nlm.nih.gov/32665951/> <https://pubmed.ncbi.nlm.nih.gov/18480428/> <https://pubmed.ncbi.nlm.nih.gov/24315502/> <https://pubmed.ncbi.nlm.nih.gov/22267763/> <https://pubmed.ncbi.nlm.nih.gov/16319891/> [https://erj.ersjournals.com/content/46/suppl\\_59/OA1473](https://erj.ersjournals.com/content/46/suppl_59/OA1473) [https://erj.ersjournals.com/content/48/suppl\\_60/PA3352](https://erj.ersjournals.com/content/48/suppl_60/PA3352)

The primary objective of this thesis is to develop a R package that provides easy and efficient access to the predetermined subject in order to conduct research that is directly related to making a significant contribution to patient care via the use of pheno- and endotype categorisation. Conducting research on cluster-analysis-based classifications should be encouraged, as they have aided physicians and academics in understanding the complicated issues associated with asthma.

Numerous studies on unsupervised learning algorithms for discovering prospective pheno- and endotypes have been conducted, but the existence of subgroups of asthma patients who do not respond well to conventional therapy indicates that certain phenotypes have not been effectively recognized and described.

The R package aims to offer researchers with fast tools for clustering patient time series data and confirming the distinction using additional metrics such as population parameter enrichment analysis, stability after random data removal, and conventional cluster stability measures.

It is also considerable to apply these data scientific techniques to other model development areas where a distinction in heterogeneous groups due to time series data fluctuations is desired.